



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

PIWI-interacting RNAs

Citation for published version:

Ozata, DM, Gainetdinov, I, Zoch, A, O'Carroll, D & Zamore, PD 2018, 'PIWI-interacting RNAs: small RNAs with big functions', *Nature Reviews Genetics*. <https://doi.org/10.1038/s41576-018-0073-3>

Digital Object Identifier (DOI):

[10.1038/s41576-018-0073-3](https://doi.org/10.1038/s41576-018-0073-3)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Reviews Genetics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



piRNAs: small RNAs with big functions

Deniz M Ozata^{1}, Ildar Gainetdinov^{1*}, Ansgar Zoch², Dónal O'Carroll² and Phillip D. Zamore¹*

¹RNA Therapeutics Institute and Howard Hughes Medical Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA;

²MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH16 4UU, UK

*These authors contributed equally to this work

Correspondence to P.D.Z

e-mail: phillip.zamore@umassmed.edu

Abstract

In animals, 21–35 nt long PIWI-interacting RNAs (piRNAs) silence transposable elements, regulate gene expression, and fight viral infection. piRNAs guide PIWI proteins to cleave target RNA, promote heterochromatin assembly, and methylate DNA.

- 5 The architecture of the piRNA pathway allows it both to provide adaptive, sequence-based immunity to rapidly evolving viruses and transposons and to regulate conserved host genes. piRNAs silence transposons in the germline of most animals, while somatic piRNA functions have been lost, gained, and lost again across evolution. Moreover, most piRNA pathway proteins are deeply conserved, but different animals employ
- 10 remarkably divergent strategies to produce piRNA precursor transcripts. Here, we discuss how a common piRNA pathway allows animals to recognize diverse targets, ranging from selfish genetic elements to genes essential for gametogenesis.

Introduction

PIWI-interacting RNAs (piRNAs) are an animal-specific class of small-silencing RNAs, distinct from microRNAs (miRNAs) and small interfering RNAs (siRNAs). piRNAs bear 2'-O-methyl modified 3' termini and guide PIWI-clade Argonautes (PIWI proteins) rather than the AGO-clade proteins which function in the miRNA and siRNA pathways (BOX 1; REF. {Aravin et al., 2006, #9748; Girard et al., 2006, #74446; Vagin et al., 2006, #48605; Lau et al., 2006, #70867; Grivna et al., 2006, #3944; Saito et al., 2006, #22846; Houwing et al., 2007, #79141; Batista et al., 2008, #91388; Das et al., 2008, #74839; Horwich et al., 2007, #59920; Saito et al., 2007, #28100; Ohara et al., 2007, #60045; Montgomery et al., 2012, #7518; Ohara et al., 2007, #60045; Kirino and Mourelatos, 2007, #70456; Horwich et al., 2007, #59920; Saito et al., 2007, #28100; Kirino and Mourelatos, 2007, #34897; Lim et al., 2015, #83471; Billi et al., 2012, #101266; Kamminga et al., 2012, #91395; Montgomery et al., 2012, #7518; Kamminga et al., 2010, #25898}).

miRNAs and siRNAs derive from double-stranded RNA precursors, but piRNAs are processed from long single-stranded precursor transcripts{Vagin et al., 2006, #48605; Houwing et al., 2007, #79141; Aravin et al., 2006, #9748; Girard et al., 2006, #74446; Brennecke et al., 2007, #38790}. The exception is nematodes, whose piRNAs (21U-RNAs) are made one-at-a-time from 25–27 nt single-stranded precursors, each transcribed from its own mini-gene{Ruby et al., 2006, #20885; Cecere et al., 2012, #21527; Gu et al., 2012, #34816}. piRNA precursors are transcribed from genomic loci known as piRNA clusters. In many arthropods, piRNA clusters correspond to large graveyards of transposon remnants{Brennecke et al., 2007, #38790; Fu et al., 2018, #18933; Kawaoka et al., 2009, #94517; Lewis et al., 2018, #102179}; in birds and mammals, piRNA clusters give rise to long non-coding RNAs (lncRNAs), which are processed into piRNAs{Aravin et al., 2006, #9748; Girard et al., 2006, #74446; Li et al.,

2013, #8343}. piRNA sequences are immensely diverse and rarely conserved among species (FIG. 1a).

In most animals, at least a subset of piRNAs defend the germline genome against transposon mobilization{Vagin et al., 2006, #48605; Aravin et al., 2007, #94199; Brennecke et al., 2007, #38790; Houwing et al., 2007, #79141; Aravin et al., 2008, #9561; Kuramochi-Miyagawa et al., 2008, #36476}. How the piRNA pathway discriminates between self and non-self transcripts remains a central question in piRNA research. This review discusses current models for piRNA cluster transcription, piRNA biogenesis, and piRNA functions in the context of the developmental challenges faced by different animals.

Discovery of piRNAs

piRNAs were first identified in the fly testis as a novel class of “long siRNAs” that silence *Stellate*, a multi-copy gene on the *Drosophila melanogaster* X-chromosome{Aravin et al., 2001, #228}. Unchecked, the *Stellate* protein crystalizes in spermatocytes, impairing male fertility{Belloni et al., 2002, #607; Bozzetti et al., 1995, #69472; Hardy et al., 1984, #47441; Livak, 1984, #17527; Livak, 1990, #81961; Meyer, 1961, #20770}.

Consequently, the Y-chromosome has amassed many copies of *Suppressor-of-Stellate*, a piRNA-producing gene derived from *Stellate* itself{Aravin et al., 2001, #228; Aravin et al., 2004, #13479}. The subsequent discovery that *flamenco* — a gene long known to repress *gypsy* family transposons — produced piRNAs rather than encoding a protein, united piRNAs with earlier genetic studies of transposon silencing and implicated the protein Piwi as central to this process{Sarot et al., 2004, #100681}.

piRNAs guide PIWI proteins in gonads of insects{Vagin et al., 2006, #48605; Saito et al., 2006, #22846}, mammals{Aravin et al., 2006, #9748; Girard et al., 2006, #74446; Lau et al., 2006, #70867; Grivna et al., 2006, #3944}, nematodes{Batista

et al., 2008, #91388; Das et al., 2008, #74839}, and fish{ Houwing et al., 2007, #79141}. To date, piRNAs and PIWI proteins have been found in the vast majority of animals, except for several species including most nematodes{Sarkies et al., 2015, #1614; Grimson et al., 2008, #14357; Mondal et al., 2018, #9707}.

5 Challenges of transposon silencing in animals

Every animal genome fights an endless war against parasitic transposable elements. Over evolutionary time, battles are won and new conflicts begin. The war is fought in the genome of the germline: transposons must integrate into the germ cell DNA to survive. Once transposons are silenced, mutations ultimately inactivate transposon-encoded
10 proteins, leading to the transposon's demise. The saga of host–transposon conflict is best understood for the *D. melanogaster* and mouse piRNA pathways, which highlight the common and distinct challenges faced by germline genomes in different animals.

How the germline is specified defines the first challenge for piRNA-based transposon defense. In many animals, including most arthropods and many chordates,
15 maternally deposited factors define primordial germline cells from which the entire germ lineage descends (reviewed in REF. {Johnson et al., 2011, #29651}). Maternal specification of the germline provides direct continuity of germ cells across generations and the opportunity for mothers to transmit information — in the form of RNA and protein — about the transposons present in the maternal genome. Indeed, PIWI
20 proteins and piRNAs are maternally deposited in insect oocytes, thus providing progeny with immunity to transposable elements{Brennecke et al., 2008, #70330; Kawaoka et al., 2011, #68142; de Vanssay et al., 2012, #4305; Le Thomas et al., 2014, #53809; Le Thomas et al., 2014, #5630; Ninova et al., 2017, #79756}. However, maternally inherited piRNAs cannot protect progeny from novel transposons present only in the
25 father. In flies, for example, when naïve mothers mate with fathers bearing genomic

insertions of the P-element transposon, the offspring are sterile because they cannot silence P-elements in their own germ cells{Kidwell and Kidwell, 1976, #47237;Rubin et al., 1982, #37378;Brennecke et al., 2008, #70330;Khurana et al., 2011, #63841}.

Many animals, including amniotes other than birds, employ a different, probably
5 ancestral mode of germline specification: somatic cells are induced to become germ cell progenitors late in development. This strategy eliminates the generational continuity of the germline, requiring the piRNA pathway to recognize transposon sequences without prior information. Moreover, the acquisition of primordial germ cells from the soma requires germline reprogramming to reset the epigenome and erase genomic
10 imprinting{Leitch et al., 2013, #72464}. In mice, germ cell reprogramming erases the DNA methylation that silences transposons, causing a burst of transposon transcription to which the piRNA pathway must respond.

Gonad anatomy and transposon life-cycle also create specific challenges for germ cells. In *D. melanogaster*, both germline stem cells and differentiating germ cells
15 contact supporting somatic cells. Several endogenous retroviruses in these somatic cells can produce infectious virions able to infect adjacent germ cells{Chalvet et al., 1999, #580}. *D. melanogaster* has evolved an abridged piRNA pathway in somatic follicle cells to counteract this threat{Sarot et al., 2004, #100681}. Mouse spermatogonial stem cells do not face such a challenge, as the active transposons in
20 mice cannot produce infectious particles{Dewannieux and Heidmann, 2005, #4985; Dewannieux et al., 2004, #15587}.

Finally, changes in chromatin during gametogenesis pose unique challenges for the restriction of transposons. For example, in mice, meiosis includes a period of transcriptional quiescence and loss of repressive chromatin marks that is followed by
25 resumption of transcription and concomitant depression of many transposon promoters{Davis et al., 2017, #78352}. The loss of transcriptional repression

necessitates continuous post-transcriptional silencing of transposon mRNAs by piRNAs throughout mouse spermatogenesis{Reuter et al., 2011, #12609;Di Giacomo et al., 2013, #51366}. The piRNA pathway provides both innate and adaptive solutions to these challenges.

5 **piRNA biogenesis**

Genomic sources of piRNAs. What defines a piRNA-producing gene and what marks its transcripts for piRNA production remains a central unsolved question in the field.

Historically, piRNA-producing loci have been called “clusters,” because they were initially defined by the high density of piRNAs mapping to them{Girard et al., 2006, 10 #74446; Grivna et al., 2006, #3944; Aravin et al., 2006, #9748; Lau et al., 2006, #70867}. In flies, piRNA precursors come from heterochromatic loci (FIG. 1b,c), whereas in mammals, piRNA clusters appear to be indistinguishable from canonical euchromatic RNA polymerase II (RNA Pol II) transcription units (FIG. 1d). “Uni-strand” clusters generate piRNA precursors by conventional, unidirectional transcription and 15 have been found in all piRNA-producing animals examined to date (FIG. 1c,d). “Dual-strand” clusters, which are convergently transcribed from both DNA strands, have been identified in dipterans{Klattenhoff et al., 2009, #38065; Mohn et al., 2014, #77489} and lepidopterans{Fu et al., 2018, #18933}, and are likely present in other arthropods (FIG. 1b).

20 In flies, piRNA clusters record a species’ history of transposon invasion, allowing piRNAs to silence the large number of active transposon families present in the *D. melanogaster* genome. In the germline, dual-strand clusters produce the majority of fly piRNAs{Brennecke et al., 2007, #38790}. Fly dual-strand clusters lack the hallmarks of canonical transcription, such as the “active” promoter mark histone H3 trimethyl lysine 4 25 (H3K4me3) and the use of standard RNA signal sequences to remove introns and

terminate transcription. Dual-strand clusters, by their nature, produce sense and anti-sense piRNAs regardless of transposon orientation{Li et al., 2009, #70761; Malone et al., 2009, #67411}. The current model for piRNA production from dual-strand clusters seeks to explain the findings that their transcription requires both the transcriptionally repressive histone H3 chromatin mark, trimethyl lysine 9 (H3K9me3), and the piRNA-guided, transcriptional silencing protein Piwi{Moshkovich and Lei, 2010, #63298; Rangan et al., 2011, #93338}.

Dual-strand clusters make piRNA precursor RNAs via non-canonical, transcription facilitated by the germline-specific, H3K9me3-binding protein Rhino, a variant of HP1{Klattenhoff et al., 2009, #38065; Mohn et al., 2014, #77489; Chen et al., 2016, #40928; Pane et al., 2011, #61432; Andersen et al., 2017, #7306}. Together with Deadlock (Del) and Cutoff (Cuff), Rhino bypasses the need for promoter sequences. Binding of Rhino to H3K9me3 tethers the germline-specific TFIIA-L paralog, Moonshiner, along both strands of the piRNA cluster DNA. Moonshiner, in turn, forms an alternative TFIIA pre-initiation complex with TATA box-binding protein-related factor 2, allowing RNA Pol II to initiate dual-strand cluster transcription “incoherently,” i.e., from many sites and on both DNA strands{Andersen et al., 2017, #7306}. Thus, dual-strand piRNA cluster transcription reflects the occupancy of chromatin by Rhino rather than specific DNA regulatory sequences such as promoters (FIG. 1b).

With Cuff and Del, Rhino also represses the splicing of dual-strand cluster transcripts and the use of canonical cleavage and polyadenylation sequence motifs{Zhang et al., 2014, #20677; Chen et al., 2016, #40928} (FIG. 1b). Cuff has been proposed to compete with cap-binding proteins that promote splicing, and, together with UAP56 and THO-complex proteins, to send cluster transcripts to the piRNA-producing machinery present in nuage, a specialized perinuclear structure unique to germ cells{Hur et al., 2016, #60320; Zhang et al., 2012, #59323}. Alas, Rhino, Cuff, Del, and

Moonshiner, so central to the identity of fly dual-strand piRNA clusters, are rapidly evolving and not found outside of Drosophilids{Vermaak et al., 2005, #93200; Fu et al., 2018, #18933; Parhad et al., 2017, #41716}.

In the somatic follicle cells that support fly oogenesis, uni-strand clusters alone
5 prevent endogenous retroviruses from infecting adjacent germ cells{Brennecke et al., 2007, #38790; Li et al., 2009, #70761; Malone et al., 2009, #67411}. The largest somatic piRNA cluster, *flamenco*, resides in the pericentromeric heterochromatin of the X chromosome, yet is conventionally transcribed to produce a long precursor transcript that generates piRNAs{Brennecke et al., 2007, #38790; Li et al., 2009, #70761; Malone
10 et al., 2009, #67411; Mevel-Ninio et al., 2007, #6139; Prud'homme et al., 1995, #92872; Sarot et al., 2004, #100681}. *flamenco* contains many antisense transposon sequences, allowing it to directly produce piRNAs that target transposon mRNAs. The transcription factor Cubitus interruptus drives *flamenco* transcription{Goriaux et al., 2014, #40466} (FIG. 1c). Conservation of *flamenco* and its Cubitus interruptus-binding site across
15 Drosophilids suggest that the cluster arose recently in arthropod evolution, but before Drosophilid speciation (≥ 50 million years ago; REFS {Goriaux et al., 2014, #40466; Malone et al., 2009, #67411; Chirn et al., 2015, #5630; Zanni et al., 2013, #20103}). Other arthropods likely possess uni-strand piRNA clusters that are evolutionarily unrelated to *flamenco*{Fu et al., 2018, #18933}, and we do not yet know whether they
20 share common strategies for designating the transcripts of uni-strand clusters as piRNA precursors.

What features distinguish *flamenco* from other conventional long non-coding RNAs that do not produce piRNAs? Alternative splicing of *flamenco* has been proposed to promote the binding of UAP56 and exportins, proteins that help ferry *flamenco* RNA
25 from the nucleus to cytoplasmic piRNA processing sites{Dennis et al., 2016, #44032; Handler et al., 2013, #62852; Muerdter et al., 2013, #86013} (FIG. 1c). However,

UAP56 and exportins also transport transcripts that produce no piRNAs. A recent study reported that the protein Yb binds *cis*-acting RNA elements in the *flamenco* transcript in the cytoplasm, triggering its processing into piRNAs{Ishizu et al., 2015, #13056; Homolka et al., 2015, #56655; Pandey et al., 2017, #44692}. Given that *flamenco* is an evolutionarily young innovation{Goriaux et al., 2014, #40466; Malone et al., 2009, #67411;Fu et al., 2018, #18933}, the mechanism in which an RNA-binding protein recognizes specific sequences in a piRNA precursor transcript maybe unique among *Drosophilids*.

Although dual-strand piRNA clusters have not been identified outside arthropods, uni-strand clusters may play a role in mammalian transposon silencing during fetal spermatogenesis. In the mouse testis, PIWI proteins appear around the thirteenth day of embryonic development{Aravin et al., 2008, #9561; Kuramochi-Miyagawa et al., 2008, #36476}. Thereafter, piRNA production persists throughout spermatogenesis{Girard et al., 2006, #74446; Lau et al., 2006, #70867; Aravin et al., 2006, #9748; Aravin et al., 2007, #94199; Aravin et al., 2008, #9561; Li et al., 2013, #8343}, except in the leptotene and zygotene stages of meiosis, during which PIWI proteins have not been detected{Di Giacomo et al., 2013, #51366}. Among the loci that produce transposon silencing piRNAs in the fetal mouse testes are two uni-strand clusters that, like *flamenco* in flies, produce piRNAs mainly antisense to transposon mRNAs{Aravin et al., 2008, #9561}.

In adult male mice, piRNAs in germs cells at stages before pachytene (*pre-pachytene* piRNAs) include transposon-silencing piRNAs, but the majority of piRNAs derive from the coding and 3' untranslated regions (3' UTRs) of hundreds of mRNAs{Li et al., 2013, #8343; Gainetdinov et al., 2018, #66022}. Such 3' UTR piRNAs have been found in the somatic and germline tissues of just a few arthropods, including the follicle cells of the fly ovary{Robine et al., 2009, #98156;Lewis et al., 2018, #102179}. Why

some mRNAs make piRNAs and what purpose these sense piRNAs serve remains mysterious. Perhaps 3' UTR piRNAs act in *trans* to regulate partially complementary mRNAs, or maybe they simply reflect co-option of the piRNA pathway to target some mRNAs for destruction.

5 *Pachytene* piRNAs begin to accumulate in spermatocytes at the pachytene stage of meiosis, representing ~95% of all piRNAs in the adult mouse testis{Aravin et al., 2006, #9748; Girard et al., 2006, #74446}. *Pachytene* piRNA precursors are transcribed from ~100 standard lncRNA genes that are depleted of transposons compared to the rest of the genome{Li et al., 2013, #8343}. The transcription factor A-
10 MYB (MYBL1) coordinately initiates transcription of these loci as well as numerous piRNA biogenesis components, including MIWI (PIWIL1), MILI (PIWIL2), and VASA (DDX4; REF. {Bolcun-Filas et al., 2011, #43995; Li et al., 2013, #8343}) (FIG. 1d). *Pachytene* piRNA loci are often divergently transcribed from a central promoter, and some of these produce a piRNA precursor transcript from one arm and an mRNA or
15 lncRNA from the other. Again, why pachytene piRNA precursor transcripts make piRNAs, while other A-MYB-regulated mRNAs and lncRNAs do not, is currently unknown.

Caenorhabditis elegans, unlike most nematodes, retains a piRNA pathway, albeit highly evolutionarily derived{Sarkies et al., 2015, #1614}. For historical reasons, *C.*
20 *elegans* piRNAs are called 21U-RNAs, reflecting their length and first nucleotide bias, and many aspects of *C. elegans* piRNA production and function have not yet been observed outside of roundworms. *C. elegans* presents an exception to the general mechanism of piRNA production from long precursor RNAs. *C. elegans* Type I 21U-RNAs are produced from ~12,000 dedicated mini-genes controlled by the same set of
25 proteins, including the Forkhead family transcription factor FKH and a MYB-like transcription factor SNPC-4, which is assisted by the nuclear protein PRDE-1 (FIG. 1e,

REF. {Weick et al., 2014, #68696; Kasper et al., 2014, #23755; Cecere et al., 2012, #21527}). Each Type I mini-gene generates a 7-methylguanosine-capped piRNA precursor transcript just 25–27 nt long{Gu et al., 2012, #34816; Weick et al., 2014, #68696; Ruby et al., 2006, #20885} (FIG 1e). In contrast, the Type II 21U-RNAs are
5 generated at the transcription start sites of conventional protein-coding genes and other RNA Pol II transcripts{Gu et al., 2012, #34816}. RNA Pol II pausing or premature termination is hypothesized to produce both types of 21U-RNAs{Gu et al., 2012, #34816; Beltran et al., 2018, #9000}, which are initially 25–27 nt long and subsequently processed by an as yet unknown mechanism that removes the 7-methyl guanosine cap
10 and the first two nucleotides of the precursor, establishing U as the first nucleotide (FIG. 2).

Making piRNA 5' ends. From flies to mice to worms, piRNA precursor transcripts begin with a 7-methylguanosine cap, yet piRNAs start with a 5' monophosphate. The first step in committing an RNA to produce piRNAs appears to be endonucleolytic cleavage that
15 generates the monophosphorylated end (FIG. 2 and 3) required for PIWI protein binding to RNA{Kawaoka et al., 2011, #25706; Cora et al., 2014, #45361; Wang et al., 2014, #77252; Matsumoto et al., 2016, #4864}. The requirement for a 5' monophosphate licenses piRNA precursors: only long, single-stranded, 5' monophosphorylated RNAs can enter the piRNA pathway{Wang et al., 2014, #77252; Mohn et al., 2015, #6778;
20 Han et al., 2015, #29200; Wang et al., 2015, #21998}.

In most animals, current evidence suggests that two pathways make piRNA 5' ends. In the first, slicing of long precursor transcripts by piRNA-guided PIWI proteins initiates the production of piRNAs via a process known as the ping-pong cycle (FIG. 3; REF. {Brennecke et al., 2007, #38790; Gunawardane et al., 2007, #65138}). The ping-
25 pong pathway begins when a PIWI protein, guided by an ***initiator*** piRNA, cleaves a

complementary target transcript to generate a **pre-pre-piRNA** with a monophosphorylated 5' end. In many animals, initiator piRNAs are maternally inherited. Binding of a PIWI protein to the pre-pre-piRNA commits the RNA to produce a **responder** piRNA from its 5' end. The establishment of the responder piRNA 3' end is initiated by the endonuclease in the second pathway (see below). In many animals, the intermediate product of this process—a **pre-piRNA** bound to the PIWI protein—is longer than a piRNA and must be trimmed to generate the mature responder piRNA. Because all Argonaute proteins, including PIWI proteins, slice their targets between nucleotides 10 and 11 of their guide, the first ten nucleotides of the responder piRNA are complementary to the first ten nucleotides of the initiator piRNA directing the cut (FIG. 3). The new responder piRNA can itself act as an initiator piRNA, producing a new responder piRNA identical to the original initiator piRNA. Thus, the ping-pong pathway functions as an amplification loop, limited only by available piRNA precursor substrates (FIG. 3).

In the second pathway, a piRNA-independent endonuclease in a complex of proteins on the mitochondrial outer membrane establishes the 3' of the responder pre-piRNA. The same complex of proteins converts the remaining 3' section of the pre-pre-piRNA into a string of tail-to-head, “phased” **trailing** pre-piRNAs {Mohn et al., 2015, #6778; Han et al., 2015, #29200}. The two pathways collaborate: the ping-pong pathway fragments long piRNA precursor transcripts, creating 5' monophosphorylated pre-pre-piRNA that provide entry points for the production of responder and trailing piRNAs by the phased piRNA pathway (FIG. 3).

Before the discovery that initiator piRNAs are upstream of responder and trailing piRNAs, terms “primary” for trailing piRNAs and “secondary” for initiator and responder piRNAs were used. We propose to replace these historical terms with the more intuitive names initiator/responder, ping-pong piRNAs and trailing, phased piRNAs.

Which pathway creates a piRNA influences its nucleotide sequence. In the phased piRNA pathway, trailing pre-piRNAs often start with uridine (1U bias; FIG. 3; REF. {Aravin et al., 2006, #9748; Brennecke et al., 2007, #38790}). This 1U bias likely reflects the specificity of the endonuclease — thought to be Zucchini (called PLD6 in mammals) — that generates the ends of phased pre-piRNAs{Haase et al., 2010, #5392; Ipsaro et al., 2012, #70084; Nishimasu et al., 2012, #101700}. In contrast, piRNAs generated by the ping-pong pathway bear a characteristic adenine at position 10 (10A; REF. {Brennecke et al., 2007, #38790; Gunawardane et al., 2007, #65138; Kawaoka et al., 2009, #94517; Houwing et al., 2008, #27066}). Although base pairing between 1U-biased initiator piRNAs and target transcripts is a logical explanation for the 10A in responder piRNAs, it is not the actual source of the 10A signature, because the structure of Argonaute does not allow the first nucleotide of a guide RNA (g1) to base pair with corresponding target nucleotide (t1; REF. {Haley and Zamore, 2004, #789; Ma et al., 2005, #26348; Parker et al., 2005, #818; Wang et al., 2009, #39503; Frank et al., 2010, #58056; Boland et al., 2011, #49469; Kawaoka et al., 2011, #25706; Elkayam et al., 2012, #33378; Schirle and MacRae, 2012, #21385; Schirle et al., 2014, #6769; Cora et al., 2014, #45361; Matsumoto et al., 2016, #4864}). Indeed, structural data and analysis of the target preferences of PIWI proteins from various animals reveals that many possess an intrinsic affinity for a t1 adenine, regardless of the identity of the g1 nucleotide{Wang et al., 2014, #77252; Matsumoto et al., 2015, #14343}. When an initiator piRNA directs a PIWI protein (e.g., fly Aub) to bind and slice a target, its preference for t1A targets generates a responder piRNA with g10A. That is, the t1A of the target RNA becomes g10A of the responder piRNA. The responder piRNA with g10A guides PIWI-catalyzed slicing of targets with t10U because of complementary pairing with g10A. Slicing converts t10U to g1U in the resulting responder piRNA. Consequently, the preference of PIWI proteins for t1A targets is one of the sources of

the g1U bias of its piRNA guides. The preference of PIWI proteins for t1 adenine parallels that of miRNAs, which repress targets more efficiently when they bear t1 adenine, because miRNA-binding Argonautes contain a t1A “pocket” that reads the target nucleotide identity{Lewis et al., 2005, #8729; Grimson et al., 2007, #84521; Nielsen et al., 2007, #4899; Baek et al., 2008, #23842; Selbach et al., 2008, #85967; Schirle et al., 2015, #61513}.

Ping-pong amplification increases the abundance of pre-existing piRNAs, while the phased piRNA pathway expands the diversity of piRNA sequences by spreading piRNA production 5'-to-3' downstream of the cut directed by the initiator piRNA. Recent studies of flies and mice{Mohn et al., 2015, #6778; Han et al., 2015, #29200; Senti et al., 2015, #7485; Wang et al., 2015, #21998; Homolka et al., 2015, #56655; Yang et al., 2016, #30387}, and data from an evolutionarily broad range of non-model species{Gainetdinov et al., 2018, #66022} suggest that in most animals the ping-pong and phased piRNA pathways collaborate to make complex populations of piRNAs.

Polishing piRNA 3' ends. Trimming and 2'-O-methylation of pre-piRNA 3' ends concludes piRNA biogenesis{Kawaoka et al., 2011, #25706; Tang et al., 2016, #19230; Izumi et al., 2016, #4722; Hayashi et al., 2016, #59204; Horwich et al., 2007, #59920; Saito et al., 2007, #28100; Kirino and Mourelatos, 2007, #34897; Lim et al., 2015, #83471}. piRNA 3' terminal 2'-O-methylation has been hypothesized to enhance small RNA stability, by protecting the piRNA from non-templated nucleotide addition and 3'-to-5' exonucleases{Kamminga et al., 2010, #25898; Lim et al., 2015, #83471}. How piRNA trimming supports piRNA function is not known. In many animals, the 25–50 nt long pre-piRNAs require extensive 3' trimming to generate functional piRNAs{Izumi et al., 2016, #4722; Ding et al., 2017, #33444; Zhang et al., 2017, #52582; Tang et al., 2016, #19230; Nishimura et al., 2018, #44552; Gainetdinov et al., 2018, #66022}. Studies in silkworm and mouse identified Trimmer/PNLDC1 as the exonuclease that trims pre-

piRNAs{Izumi et al., 2016, #4722; Ding et al., 2017, #33444; Zhang et al., 2017, #52582; Nishimura et al., 2018, #44552}; its ortholog PARN-1 trims *C. elegans* pre-piRNAs{Tang et al., 2016, #19230}. In mice and worms, Trimmer is required for fertility, but in flies, trimming is largely dispensable, likely because fly pre-piRNAs are often no longer than mature piRNAs{Han et al., 2015, #29200}. Thus, it is not surprising that the mechanism by which flies shorten their pre-piRNA 3' ends is evolutionarily atypical. *D. melanogaster* and the rest of the Brachycera suborder of Diptera lost orthologs of both PNLDC1 and PARN-1 ~270 million years ago{Hedges et al., 2015, #44854; Hayashi et al., 2016, #59204} and use the miRNA-trimming exonuclease Nibbler to resect piRNAs{Han et al., 2011, #98377; Liu et al., 2011, #19273; Feltzin et al., 2015, #33454; Hayashi et al., 2016, #59204}.

piRNA 3' ends are 2'-O-methylated by an S-adenosylmethionine-dependent methyltransferase (Hen1 in flies; HENMT1 in mice, HENN-1 in worms; REF. {Vagin et al., 2006, #48605; Ohara et al., 2007, #60045; Kirino and Mourelatos, 2007, #70456; Houwing et al., 2007, #79141; Horwich et al., 2007, #59920; Saito et al., 2007, #28100; Kirino and Mourelatos, 2007, #34897; Lim et al., 2015, #83471; Billi et al., 2012, #101266; Kamminga et al., 2012, #91395; Montgomery et al., 2012, #7518; Kamminga et al., 2010, #25898}). HEN1 was first discovered in plants, where it modifies siRNAs and miRNAs{Li et al., 2005, #96284; Yu et al., 2005, #65268}. Consistent with a role for 2'-O-methylation in stabilizing piRNAs, PIWI proteins bind more tightly to 2'-O-methyl-modified 3' ends{Tian et al., 2011, #24490; Simon et al., 2011, #70780; Zeng et al., 2011, #76288; Matsumoto et al., 2016, #4864}.

piRNAs are made in specialized cytoplasmic compartments. Most piRNA pathway proteins localize to specific cytoplasmic compartments, including *nuage* in animal germ cells, Yb bodies in the somatic ovarian follicle cells of flies, and the mitochondrial outer

membrane in all phased piRNA-producing cells (BOX 2). The enrichment of the piRNA machinery in these subcellular structures may serve to increase the local concentration of specific proteins or protect piRNA precursors from housekeeping nucleases.

Compartmentalization may also prevent mRNAs and lncRNAs from entering the piRNA pathway. Supporting the idea that piRNA precursor transcripts are shunted to sites of piRNA production, nuclear-localized piRNA pathway proteins can be found opposite nuage-localized piRNA pathway proteins across the channel of a single nuclear pore in fly germline cells{Zhang et al., 2012, #59323}, suggesting that piRNA precursor transcripts are exported from the nucleus directly into nuage, unlike other cytoplasmic RNAs.

Factors that initiate piRNA biogenesis by generating pre-pre-piRNAs are found in nuage (BOX 2, Table 1), for example fly Aub, Ago3, and Vasa{Liang et al., 1994, #24175; Harris and Macdonald, 2001, #299; Findley et al., 2003, #17179; Brennecke et al., 2007, #38790; Lim and Kai, 2007, #84528; Zhang et al., 2011, #79304}. Male mouse fetal germ cells contain two types of nuage. MILI and TDRD1 are found in nuage localized between clusters of mitochondria{Chuma et al., 2006, #46434; Aravin et al., 2009, #72503}, i.e., the classical intermitochondrial cement {Eddy, 1975, #102610}, whereas MIWI2, MAEL and TDRD9 are in perinuclear nuage granules{Aravin et al., 2009, #72503; Shoji et al., 2009, #21684}. In contrast, factors such as Zucchini/PLD6, Gasz/GASZ, and Papi/TDRKH, which are required to generate responder pre-piRNA 3' ends or produce both ends of trailing pre-piRNAs, localize to the outer membrane of mitochondria{Choi et al., 2006, #18782; Wang et al., 2007, #81096; Saito et al., 2010, #47914; Watanabe et al., 2011, #9704; Huang et al., 2011, #46149; Handler et al., 2013, #62852; Honda et al., 2013, #59984; Saxe et al., 2013, #76672}. The enzyme PNLDC1/Trimmer carrying out 3'-to-5' trimming of responder and trailing pre-piRNAs is also believed to be located at the same site where these piRNAs are produced, i.e., on

the outer membrane of mitochondria{Izumi et al., 2016, #4722}. Conversely, in the evolutionarily derived architecture of the fly piRNA pathway, the piRNA trimming enzyme Nibbler is found in the perinuclear nuage and does not trim trailing piRNAs{Hayashi et al., 2016, #59204}. In fly ovarian somatic follicle cells, granules
5 called Yb bodies have also been implicated in non-ping-pong piRNA biogenesis{Szakmary et al., 2009, #24009; Saito et al., 2010, #47914}. Rigorous proof that sites with high piRNA pathway protein concentrations participate in piRNA production continues to be elusive, but it is striking that cellular compartments containing proteins that act sequentially in piRNA biogenesis — e.g., nuage and
10 mitochondria — are frequently near one another{Szakmary et al., 2009, #24009; Saito et al., 2010, #47914; Eddy, 1974, #27970; Eddy, 1975, #102610; Aravin et al., 2009, #72503; Shoji et al., 2009, #21684}. In support of nuage serving as a piRNA factory, artificially tethering nuage proteins to an RNA triggers its processing into piRNAs{Pandey et al., 2017, #44692; Rogers et al., 2017, #56941}

15 Tudor domains — four-stranded β -barrels — scaffold the assembly of complex cellular machines by binding symmetrically dimethylated arginine (sDMAs), a modification found on PIWI proteins. Tudor-domain proteins have been proposed to coordinate nuage assembly and tether PIWI proteins to the outer face of mitochondria{Izumi et al., 2016, #4722; Siomi et al., 2010, #63382; Chen et al., 2009,
20 #19756; Nishida et al., 2009, #96922; Reuter et al., 2009, #86589; Wang et al., 2009, #63594; Vagin et al., 2009, #7334; Kirino et al., 2010, #57671; Huang et al., 2011, #68875; Patil and Kai, 2010, #101768; Anand and Kai, 2012, #12689; Webster et al., 2015, #37148; Sato et al., 2015, #29697; Nishida et al., 2018, #57554}.

In flies, Tudor-domain proteins are also required for heterotypic ping-pong
25 between the fly PIWI proteins Aub and Ago3{Zhang et al., 2011, #79304; Webster et al., 2015, #37148}. Heterotypic Aub:Ago3 ping-pong drives the production of antisense

piRNAs that direct Aub to bind transposon mRNAs in the cytoplasm and Piwi — the third fly PIWI protein — to bind nascent transposon transcripts in the nucleus. The antisense bias of fly piRNAs suggests that $k_{cat} (Ago3) \gg k_{cat} (Aub)$. In an animal inheriting antisense, Aub-bound piRNAs, a greater catalytic efficiency for Ago3 would ensure that heterotypic Aub:Ago3 ping-pong generates an excess of both antisense responder and trailing piRNAs. The Tudor-domain protein Krimper promotes heterotypic Aub–Ago3 ping-pong{Webster et al., 2015, #37148}, while the Tudor-domain protein Qin thwarts futile homotypic Aub–Aub ping-pong{Zhang et al., 2011, #79304}; together, Krimper and Qin ensure that the ping-pong cycle favors the production of piRNAs antisense to transposon mRNAs. In contrast, mouse piRNAs are amplified by homotypic MILI–MILI ping-pong{De Fazio et al., 2011, #91107}, which appears to suffice for post-transcriptional control of transposons.

piRNA functions

Studies of animals from humans to hydra suggest that silencing transposons in the germline is the ancestral function of piRNAs{Aravin et al., 2008, #9561; Juliano et al., 2014, #75805; Roovers et al., 2015, #9960; Praher et al., 2017, #77346; Gainetdinov et al., 2017, #55062; Lewis et al., 2018, #102179}. Mosquitos also use piRNAs to fight viruses in the soma{Morazzani et al., 2012, #90714; Schnettler et al., 2013, #90813; Miesen et al., 2015, #47976}. However, many piRNAs, particularly in the mammalian testis, correspond to unique genomic sequences unrelated to transposable elements{Aravin et al., 2006, #9748; Girard et al., 2006, #74446}. Although understanding the function of these evolutionarily younger, non-transposon piRNAs remains technically and intellectually challenging, accumulating evidence suggests that they regulate expression of host mRNAs.

Transposon silencing. Transposons pose multiple threats to the genome. Their presence at multiple genomic sites promotes illegitimate recombination, their replication can generate double-stranded DNA breaks, their insertion in new sites can disrupt coding sequences, and their promoters can drive aberrant expression of neighboring genes{Goodier and Kazazian, 2008, #52493; Zamudio et al., 2015, #34605; Davis et al., 2017, #78352; Vasiliauskaitė et al., 2018, #25283}. Nearly all animals rely on piRNAs to defend the germline genome from transposon expression. Arthropods and mollusks also use piRNAs to repress transposons in the soma{Lewis et al., 2018, #102179; Jehn et al., 2018, #68330}, suggesting that the last common ancestor of Protostomia and probably all other animals produced both germline and somatic piRNAs. In the evolutionarily exceptional instance of the *Drosophila* genus, piRNAs also help maintain telomeres{Savitsky et al., 2006, #964; Khurana et al., 2010, #20966; Radion et al., 2018, #59074}. Flies lack telomerase and instead use telomeric retrotransposons that recursively integrate into telomeric regions to maintain chromosome ends, (reviewed in REF. {Pardue and DeBaryshe, 2008, #32429; Pardue and DeBaryshe, 2011, #82721}). piRNA-mediated silencing of these telomeric retrotransposons sustains the heterochromatin environment required to maintain a stable telomere length{Klenov et al., 2007, #37959; Radion et al., 2018, #59074}.

In worms, piRNA-directed silencing of transposons is achieved by initiating a secondary siRNA response (see the section, “Worm piRNAs distinguish self from non-self”). In other animals, piRNAs silence transposons either by repressing their transcription or by slicing (cleaving) their mRNAs (FIG. 4a). The cytoplasmic PIWI proteins Aub and Ago3 in flies, Siwi and BmAgo3 in silkworm, and MILI and MIWI in mice mediate post-transcriptional transposon silencing{Aravin et al., 2007, #94199; Brennecke et al., 2007, #38790; Gunawardane et al., 2007, #65138; Kawaoka et al., 2009, #94517; Reuter et al., 2011, #12609; De Fazio et al., 2011, #91107; Di Giacomo

et al., 2013, #51366}. The nuclear proteins Piwi in flies and MIWI2 in mice repress transposons transcriptionally{Aravin et al., 2008, #9561; Rangan et al., 2011, #93338; Wang and Elgin, 2011, #95858; Klenov et al., 2011, #44524; Sienski et al., 2012, #17594; Le Thomas et al., 2013, #32928; Rozhkov et al., 2013, #62979}.

5 Transcriptional silencing is thought to occur when piRNA-guided PIWI proteins bind nascent transposon transcripts (FIG. 4a), a model based on siRNA-directed heterochromatin formation in *Schizosaccharomyces pombe*{Verdel et al., 2004, #661}. Transcriptional repression does not require PIWI slicer activity{Darricarrère et al., 2013, #66914; De Fazio et al., 2011, #91107}. The specific mechanism of transcriptional
10 repression differs between flies and mice. In flies, Piwi promotes K9 methylation of Histone H3, a repressive chromatin mark, through recruitment of Eggless/dSetdb1 by the Piwi-interacting mediator proteins Asterix and Panoramix{Rangan et al., 2011, #93338; Wang and Elgin, 2011, #95858; Klenov et al., 2011, #44524; Sienski et al., 2012, #17594; Le Thomas et al., 2013, #32928; Rozhkov et al., 2013, #62979; Sienski
15 et al., 2015, #44720; Yu et al., 2015, #49989; Ohtani et al., 2013, #99877; Muerdter et al., 2013, #86013; Iwasaki et al., 2016, #75201}. At dual-strand piRNA clusters and potentially at transposons themselves, the presence of H3K9me3 promotes Rhino-dependent non-canonical transcription but blocks the production of functional, spliced transposon mRNA{Andersen et al., 2017, #7306; Teixeira et al., 2017, #19643}. Thus,
20 Piwi-dependent repression of transposons in flies is not formally an example of transcriptional silencing, but rather reflects a change in the mode of transcription from producing protein-coding mRNAs to generating multiple, unspliced piRNA precursor RNAs from both genomic strands{Andersen et al., 2017, #7306; Teixeira et al., 2017, #19643}. Whether the lessons learned from *D. melanogaster* reveal evolutionarily
25 conserved principles for piRNA production is not known. Many piRNA pathway proteins

— including Rhi, Cuff, Del, and Moonshiner — are poorly conserved, lacking identifiable homologs in most other arthropods, let alone mice (Table 1).

In contrast, piRNA-dependent transcriptional silencing in mouse fetal gonocytes directs both DNA and H3K9me3 histone methylation{Carmell et al., 2007, #88633; Aravin et al., 2007, #9618; Aravin et al., 2008, #9561; Kuramochi-Miyagawa et al., 2008, #36476; Pezic et al., 2014, #67119; Molaro et al., 2014, #59995; Manakov et al., 2015, #10994; Nagamori et al., 2015, #87445; Kojima-Kita et al., 2016, #17406; Vasiliauskaitė et al., 2017, #55754; Barau et al., 2016, #58054}. Both DNA and histone H3K9me3 methylation are targeted by the piRNA pathway to evolutionarily young copies of transposons{Pezic et al., 2014, #67119; Molaro et al., 2014, #59995}. In muroid rodents, the promoters of young transposons are methylated by a dedicated DNA methyltransferase DNMT3C, encoded by a Muroidea-specific duplication of *Dnmt3b*{Barau et al., 2016, #58054; Jain et al., 2017, #39248}. How the mouse nuclear PIWI protein, MIWI2, recruits the chromatin and DNA methylation machinery to transposon genomic sequences is unknown.

In mouse male germ cells, the dramatic changes in the chromatin and transcriptional landscape during meiotic and post-meiotic stages make piRNA-guided post-transcriptional control of transposon mRNAs indispensable{Reuter et al., 2011, #12609; Di Giacomo et al., 2013, #51366}. In addition to the piRNA pathway, other repressive mechanisms silence transposons at some stages of mouse spermatogenesis{Di Giacomo et al., 2013, #51366; Di Giacomo et al., 2014, #30094}. For example, piRNA-independent H3K9me2 methylation is necessary and sufficient to silence LINE1 (long interspersed nuclear elements) transposons prior to the onset of meiosis{Di Giacomo et al., 2013, #51366; Di Giacomo et al., 2014, #30094}.

Adaptive and innate features of piRNA-directed transposon silencing. The piRNA pathway provides features of both innate and adaptive immunity against transposons.

For example, maternally deposited initiator piRNAs in *D. melanogaster* act as ‘pattern recognition receptors’ that recognize transposon sequences and respond by amplifying piRNAs specific to the threat. Indeed, a lack of innate immunity — protective maternal piRNAs — explains hybrid dysgenesis{Brennecke et al., 2008, #70330}, a phenomenon in which a transposon-carrying male mated to a naïve female produces sterile offspring, because the female cannot deposit the relevant piRNAs in her oocytes. These offspring remain sterile for most of their adult life, until adaptive piRNA-mediated immunity is reestablished in germline stem cells when the invading transposon integrates into a piRNA-producing locus{Khurana et al., 2011, #63841; Grentzinger et al., 2012, #47717}. These novel transposon insertions provide a record of the invasion by updating the piRNA cluster, and this new information immunizes future generations to the new threat.

The mouse germline is induced from somatic cells, and maternal deposition is unlikely to supply initiator piRNAs to jump-start piRNA production. However, *flamenco*-like, uni-strand piRNA-producing loci, rich in antisense transposon insertions, may provide innate memory. Broadly speaking, these uni-strand piRNA clusters may allow the piRNA machinery to recognize both “known” and — probably through partial complementarity — novel invaders. Such innate piRNAs could then trigger amplification of relevant antisense piRNAs{Aravin et al., 2008, #9561; De Fazio et al., 2011, #91107}. Moreover, the production of phased trailing piRNAs from sequences downstream of the initiator and responder piRNAs could provide adaptive immunity by favoring piRNA production from those sequences most closely related to the novel transposon. It is conceivable that these fetal, uni-strand piRNA-producing loci can accumulate transposon insertions to expand the innate memory of possible threats. Whether the A-MYB-regulated, pachytene piRNA-producing loci expressed in adult mice play a similar role remains unknown. However, pachytene piRNA-producing loci abide by conventional transcriptional rules, suggesting that unlike fly dual-strand piRNA clusters,

they are likely to be disrupted by new transposon insertions{Li et al., 2013, #8343}. This may explain why A-MYB-regulated, pachytene piRNA clusters are depleted of transposons compared to the rest of the genome{Aravin et al., 2006, #9748}.

Worm piRNAs distinguish self from non-self. *C. elegans* possesses a complex

5 system of small RNA pathways. Several studies have proposed that worm piRNAs possess broad targeting capacity, potentially recognizing any transcript present in the germline{Lee et al., 2012, #25087; Bagijn et al., 2012, #33887; Shen et al., 2018, #83072; Zhang et al., 2018, #21164}. This targeting flexibility may allow piRNAs to recognize and silence non-self transcripts such as transgenes and new transposon
10 insertions. Two models explain how self transcripts can be spared silencing (FIG. 4b): first, Argonaute CSR-1 may maintain both the transcription and stability of endogenous mRNA{Claycomb et al., 2009, #43944; Shirayama et al., 2012, #23817; Wedeles et al., 2013, #25645; Seth et al., 2013, #31627}; second, germline-expressed self transcripts may contain specific sequences conferring resistance to piRNA silencing{Zhang et al.,
15 2018, #21164}.

Unlike cytoplasmic PIWI proteins in other animals, the slicer activity of the worm PIWI protein, PRG-1, is dispensable for target silencing{Bagijn et al., 2012, #33887; Lee et al., 2012, #25087}. Instead, piRNAs induce the synthesis of secondary siRNAs on the target transcript by RNA-dependent RNA polymerase (RdRP; REFS { _Lee et al., 2012,
20 #25087; Bagijn et al., 2012, #33887; Ashe et al., 2012, #20333; Luteijn et al., 2012, #99823; Shirayama et al., 2012, #23817; Buckley et al., 2012, #76116; de Albuquerque et al., 2015, #25867} (FIG. 4b). RdRP-mediated amplification of the silencing signal is conceptually analogous to ping-pong amplification in other animals. The secondary siRNA response, RNA-induced epigenetic silencing (RNAe), can be inherited. piRNA-guided PRG-1 initiates RNAe, but other factors maintain the silencing for
25 generations{Lee et al., 2012, #25087; Bagijn et al., 2012, #33887; Ashe et al., 2012,

#20333; Luteijn et al., 2012, #99823; Shirayama et al., 2012, #23817; Buckley et al., 2012, #76116; de Albuquerque et al., 2015, #25867}.

Viral defense. Anti-viral defense in somatic tissues is typically ascribed to siRNAs.

However, some invertebrates use piRNAs to tackle viral infection in the

5 soma{Morazzani et al., 2012, #90714; Schnettler et al., 2013, #90813; Miesen et al., 2015, #47976; Lewis et al., 2018, #102179}. Mosquitoes appear to fight RNA viruses using the ping-pong pathway: two mosquito PIWI proteins — Piwi5 and Ago3 — participate in heterotypic ping-pong, consuming viral (+) and (–) strand RNAs to produce piRNAs{Miesen et al., 2015, #47976} (FIG. 4c). Genomic viral integrations
10 acting as piRNA-producing loci probably allow the mosquito piRNA pathway to recognize viral RNA by initiating the ping-pong cycle{Whitfield et al., 2017, #27135;Palatini et al., 2017, #42170}. How the piRNA pathway recognizes and tackles viral RNAs in other animals is currently unknown, as virus-derived piRNAs in other invertebrates show no signs of ping-pong{Lewis et al., 2018, #102179}.

15 Why do some animals mount piRNA-based antiviral responses, while others rely entirely on the siRNA-driven RNAi pathway for anti-viral defense? The use of two different classes of small silencing RNAs to fight viruses may reflect the distinct precursors that can enter the RNAi and piRNA pathways: RNAi is triggered by double-stranded RNA, while piRNAs are produced from single-stranded RNA. The two
20 pathways may target RNA from different types of viruses or stages of viral infection, boosting the overall anti-viral response. Testing these ideas remains an important challenge for the small RNA field.

Mammalian pachytene piRNAs: regulating gene expression? In mammals, fetal piRNAs silence transposons in male germ cells. In contrast, the most abundant piRNA
25 population in mammals, the pachytene piRNAs are depleted of transposon

sequences{Aravin et al., 2006, #9748;Girard et al., 2006, #74446}. Each spermatocyte cell contains >5 million pachytene piRNA molecules{Gainetdinov et al., 2018, #66022}. Until recently, pachytene piRNAs had not been formally proved to have a function, but a recent study reports compromised sperm function in mice lacking a major piRNA-producing locus on chromosome 6 (REF. {Wu et al., 2018, #50308}). However, the regulatory targets of pachytene piRNAs are not obvious, as >80% of pachytene piRNAs map only to the loci producing them{Aravin et al., 2006, #9748;Girard et al., 2006, #74446;Li et al., 2013, #8343}. Thus, no consensus model for how pachytene piRNAs ensure normal spermatogenesis has been established. One study reported that pachytene piRNAs guide PIWI proteins to destabilize their mRNA targets via miRNA-like mechanism{Gou et al., 2014, #29255}, while another proposed that PIWI proteins do not use pachytene piRNAs as conventional guides and instead bind and stabilize mRNAs in the sequence-independent manner{Vourekas et al., 2012, #89447}. Two recent reports show that pachytene piRNAs regulate gene expression by guiding conventional, PIWI-dependent cleavage of targets{Goh et al., 2015, #12004; Zhang et al., 2015, #22751} (FIG. 4d). Nevertheless, the minimal overlap among the targets identified in these studies suggests that we are still far from understanding what rules govern target recognition by pachytene piRNAs.

Other functions of the piRNA pathway. When discovered, the fly PIWI protein Piwi was reported to be essential for germ stem cell regeneration{Lin and Spradling, 1997, #47144;Cox et al., 1998, #181}. Whether piRNA participate in this function of Piwi has not been directly tested. In fact, the transposon-silencing function of Piwi can be genetically separated from its role in germ stem cell maintenance by removing an N-terminal region of the protein{Klenov et al., 2011, #44524}. Roles for PIWI proteins in sustaining stem cell populations have been described outside of flies, but the molecular

mechanism of these pathways also remains unknown (reviewed in REF. {Juliano et al., 2011, #104241}).

In *D. melanogaster*, the PIWI protein Aub has been reported to play a piRNA-directed role in embryonic patterning{Rouget et al., 2010, #65375; Barckmann et al., 2015, #20335}. Transposon-derived piRNAs were identified that can pair with partially complementary target sites in the 3' UTR of *nanos* mRNA and induce its decay outside the posterior pole of the embryo, where Nanos acts to repress the anterior-posterior determinant *hunchback*{Rouget et al., 2010, #65375; Barckmann et al., 2015, #20335}. However, earlier experiments identified *cis*-acting RNA elements responsible for *nanos* translational repression that do not overlap the two piRNA binding sites{Gavis et al., 1996, #41596; Gavis et al., 1996, #99744}. Whether the two regulatory mechanisms act redundantly or additively and whether piRNA-dependent mechanisms regulate *nanos* in other Diptera remain to be determined.

Conclusions

Despite 17 years of study, the central questions posed when piRNAs were first discovered remain unanswered for most animals.

First, what defines a piRNA cluster? That is, what marks a specific genomic region to produce piRNAs? Why do fly uni-strand clusters and mouse pachytene piRNA loci produce piRNAs while other mRNA and lncRNA genes do not? In flies, discrete RNA sequence elements have been suggested to funnel conventional Pol II *flamenco* transcripts into the somatic piRNA pathway{Ishizu et al., 2015, #13056; Homolka et al., 2015, #56655; Pandey et al., 2017, #44692}, but these sequences are not deeply conserved. piRNAs direct H3K9me3 marks to fly dual-strand piRNA clusters and dispersed transposon copies. In turn, H3K9me3 binds Rhino, silencing transposons and enabling cluster transcription. But many regions of the genome replete with H3K9me3

marks neither bind Rhino nor make piRNAs. What distinguishes heterochromatic piRNA clusters from other regions of heterochromatin? Moreover, Rhino homologs have not been identified outside of Drosophilids{ Vermaak et al., 2005, #93200; Parhad et al., 2017, #41716; Fu et al., 2018, #18933}, suggesting that yet undiscovered mechanisms promote piRNA production from dual-strand clusters in other arthropods{Fu et al., 2018, #18933}. The divergence of proteins involved in piRNA precursor transcription contrasts sharply with the deep conservation of the downstream piRNA-producing machinery{Grimson et al., 2008, #14357; Klattenhoff et al., 2009, #38065; Handler et al., 2011, #15938; Cecere et al., 2012, #21527; Hayashi et al., 2016, #59204; Andersen et al., 2017, #7306}. The rapid evolution of some piRNA pathway components may reflect an “evolutionary arms race” between the host genome and the rapidly evolving targets of the piRNA pathway, i.e., transposons{Simkin et al., 2013, #46000; Parhad et al., 2017, #41716; Palmer et al., 2018, #20066}.

Second, in animals that induce the germline from somatic cells and therefore do not deposit piRNAs maternally, what enables the piRNA pathway to specifically recognize transposon sequences? Are piRNAs derived from the *flamenco*-like fetal clusters in mice{Aravin et al., 2008, #9561} sufficient to start the ping-pong cycle by cleaving transposon mRNAs, triggering subsequent transcriptional and posttranscriptional repression? Or is the piRNA pathway instructed by a yet-to-be discovered transposon-sensing system?

Third, why have the ancestral somatic functions of piRNAs been lost in many animal lineages? What drives the repeated repurposing of the piRNA pathway across different animal phyla{Sarkies et al., 2015, #1614; Lewis et al., 2018, #102179}? Because the miRNA biogenesis machinery produces small RNA guides that are highly conserved among animals and the RNAi response only targets transcripts homologous to a double-stranded RNA trigger, the target repertoires of the two pathways are limited.

In contrast, the piRNA pathway makes guides from single-stranded RNA, a substrate abundant in cells. Moreover, the intrinsic imprecision of piRNA biogenesis machinery produces enormously diverse piRNA guide sequences. This may allow the target repertoire to drift during evolution, allowing the fortuitous acquisition of new targets, whose regulation by piRNAs becomes fixed when it confers a selective advantage, driving the evolution of new piRNA functions. The recurrent emergence of piRNA functions unrelated to transposon repression suggests that novel, unexpected roles for piRNAs remain to be discovered.

References

Box 1 | Argonaute family proteins [Contains an embedded figure]

Small silencing RNAs, 21–35 nt long, bind Argonaute family proteins and silence complementary transcripts either transcriptionally or post-transcriptionally. Argonaute family proteins are classified into the AGO-clade and PIWI-clades (reviewed in REF. {Cenik and Zamore, 2011, #104332; Czech and Hannon, 2011, #55689}). Small interfering RNAs (siRNAs) and microRNAs (miRNAs) are cleaved from double-stranded RNA precursors by ribonuclease III family endonucleases and guide the ubiquitously expressed AGO proteins{Hammond et al., 2001, #235; Martinez et al., 2002, #6902; Nykanen et al., 2001, #47386; Hutvagner et al., 2001, #300; Mourelatos et al., 2002, #385; Tabara et al., 1999, Cell, 99, 123-32}. Drosha generates pre-miRNAs from miRNA transcripts{Lee et al., 2003, #576}, whereas Dicer converts pre-miRNAs into mature miRNAs and long double-stranded RNA into siRNAs{Bernstein et al., 2001, #198; Grishok et al., 2001, #231; Hutvagner et al., 2001, #300; Knight and Bass, 2001, #328}. The resulting small duplex RNAs bear 5' monophosphate and 2',3' hydroxyl, 2 nt overhanging 3' ends, the hallmarks of ribonuclease III enzyme products. Once a miRNA or siRNA duplex is loaded into an Argonaute protein, the choice of guide strand reflects the relative thermodynamic stability{Schwarz et al., 2003, #590; Khvorova et al., 2003, Cell, 115, 209-16; Aza-Blanc et al., 2003, Mol Cell, 12, 627-37; Tomari et al., 2004, #763} and first nucleotide composition of its 5' ends{Ghildiyal et al., 2010, #45353}. The “passenger strand” is eliminated by passive unwinding or is cleaved by Argonaute itself; the guide strand, whose 5' end is less tightly paired, is retained{Kim et al., 2007, #32948; Leuschner et al., 2006, #32224; Matranga et al., 2005, #52398} (see the figure, part. a). Many animal siRNAs and all animal miRNAs bear 2',3' hydroxy termini, although some arthropod siRNAs are 2'-O-methylated{Lewis et al., 2018, #102179; Pelisson et al., 2007, #83481} like piRNAs.

PIWI-clade proteins are often restricted to gonadal cells and are loaded with 21–35 nt long PIWI-interacting RNAs (piRNAs). Unlike siRNAs and miRNAs, piRNAs are processed from single-stranded RNA precursors; their processing does not require Dicer{Vagin et al., 2006, #48605; Houwing et al., 2007, #79141}. Typically, piRNAs
5 begin with uridine and possess 5' monophosphate and 2'-O-methyl 3' termini{Gunawardane et al., 2007, #65138; Kirino and Mourelatos, 2007, #94453; Kirino and Mourelatos, 2007, #70456; Montgomery et al., 2012, #7518; Ruby et al., 2006, #20885; Saito et al., 2007, #28100; Vagin et al., 2006, #48605; Horwich et al., 2007, #59920}. Both AGO and PIWI proteins contain three characteristic domains: PAZ, MID, and PIWI. The PAZ domain, residing at the amino terminus, provides a binding pocket
10 for the 3' end of guide RNAs{Lingel et al., 2004, #720; Song et al., 2003, #717}. The PAZ domain differs between AGO and PIWI proteins. For example, human AGO1 binds less well to an RNA duplex containing a 3' terminal 2'-O-methyl group{Ma et al., 2004, #718}, whereas the PAZ domains of PIWI proteins better accommodate the bulky 2'-O-methyl modification{Matsumoto et al., 2016, #4864; Simon et al., 2011, #70780; Tian et al., 2011, #24490; Zeng et al., 2011, #76288}. The 5' phosphate of the guide RNA is anchored in the MID domain{Ma et al., 2005, #26348; Parker et al., 2005, #818}. The MID domain presents the seed sequence of the guide as a helix, pre-paying the entropic cost of binding to its target{Parker et al., 2009, #100708}. Target cleavage
15 occurs in the PIWI domain, whose RNase H-like fold presents a catalytic triad, aspartate-aspartate-glutamate (DDE), that positions a divalent cation, typically Mg^{2+} to hydrolyze the phosphodiester bond linking target nucleotides t10 and t11 (REF. {Elbashir et al., 2001, #332; Elbashir et al., 2001, #199; Parker et al., 2004, #767; Schwarz et al., 2004, #668; Yuan et al., 2005, #4451}). Argonaute cleavage leaves a 3'
25 hydroxyl and 5' monophosphate{Schwarz et al., 2004, #668}, allowing the use of

chemically selective high-throughput sequencing methods to identify AGO and PIWI protein cleavage products{German et al., 2008, #48636; Addo-Quaye et al., 2008, #81104; Addo-Quaye et al., 2009, #21051; German et al., 2009, #63777} (see the figure panel b).

Box 2 | Cytoplasmic foci and piRNA biogenesis

“Nuage,” French for cloud, collectively describes membraneless electron-dense structures found in animal germ cells: intermitochondrial cement in the oocytes and early spermatogenic cells of mammals{Fawcett et al., 1970, #14855; Eddy, 1974, #27970}, chromatoid bodies in the late spermatogenic cells of mammals{Benda, 1891, #5828}, perinuclear nuage in fly nurse cells and zebrafish germ cells{Mahowald, 1971, #96025; Braat et al., 1999, #95709}, and P granules in worm germ cells{Strome and Wood, 1982, #72357; Wolf et al., 1983, #76899}. Nuage proteins depend on each other for their proper localization. In mice and flies, Vasa sits at the top of the hierarchy and is essential for the localization of all other components{Liang et al., 1994, #24175; Findley et al., 2003, #17179; Harris and Macdonald, 2001, #299; Lim and Kai, 2007, #84528; Malone et al., 2009, #67411; Li et al., 2009, #70761; Chuma et al., 2006, #46434; Aravin et al., 2009, #72503}. In *C. elegans*, the proteins PGL-1 and PGL-3 are indispensable for the formation of P granules{Updike and Strome, 2010, #99545; Hanazawa et al., 2011, #61237}.

Yb bodies are electron-dense perinuclear spots in fly ovarian somatic follicle cells{Szakmary et al., 2009, #24009}. In addition to the protein Yb, these structures contain Armitage{Olivieri et al., 2010, #35447; Saito et al., 2010, #47914} and Vreteno{Handler et al., 2011, #15938}.

Accumulating evidence suggests that the properties of nuage follow the concept of liquid–liquid phase separation (reviewed in REF. {Banani et al., 2017, #102098; Seydoux, 2018, #97871}). Worm P granules behave like liquid droplets with a viscosity higher than that of the surrounding cytoplasm{Brangwynne et al., 2009, #55787}. The human homolog of Vasa, DDX4, can assemble into phase-separated membraneless bodies both in vitro and when expressed in HeLa cells{Nott et al., 2015,

#97035}. Compellingly, the domains shown or proposed to promote phase separation are also present in nuage proteins: an intrinsically disordered region in Vasa{Nott et al., 2015, #97035} and tandem TUDOR domains enabling multivalent interactions in many nuage proteins (reviewed in REF. {Chen et al., 2011, #21038}). Like other

5 membraneless organelles, nuage contains RNA as an essential component: for example, the perinuclear nuage of fly nurse cells is lost in the absence of piRNA cluster transcription{Klattenhoff et al., 2009, #38065}. Although Fluorescence Recovery After Photobleaching (FRAP) experiments show that protein components of worm P granules and fly nurse cell nuage are mobile and exchange with the cytoplasm{Webster et al.,

10 2015, #37148;Brangwynne et al., 2009, #55787}, liquid–liquid phase separation is hypothesized to slow this exchange{Banani et al., 2017, #102098}. Higher viscosity of these structures coupled with specific protein–protein interactions may maintain the distinct content of granules by retaining some biomolecules and slowing the entry of others. Future studies should help understand how exactly membraneless structures

15 contribute to piRNA biogenesis and function.

Figure Legends

FIG. 1 | **a** | Genome size and the number of piRNA species do not correlate. **b** | Fly germline dual-strand clusters exhibit “incoherent” transcription where the H3K9me3 repressive mark is recognized by Rhino (Rhi). Rhi forms a complex with Deadlock (Del) and Cutoff (Cuff). Rhi–Del recruits Moonshiner and TATA box binding protein-Related Factor 2 to the YR elements to initiate promoter-independent transcription. Rhino–Del–Cuff ensures transcriptional elongation by repressing splicing and termination at polyadenylation signal sequences (PAS) within the clusters. Thereafter, piRNA precursor transcripts are routed to nuage by UAP56. **c** | The uni-strand *flamenco* cluster resides in heterochromatin but is conventionally transcribed from a promoter element recognized by the protein Cubitus interruptus (Ci). Splicing of piRNA precursor transcripts generates several isoforms that are shuttled to cytoplasmic Yb bodies by UAP56 and the exportins Nxf1 and Nxt1. **d** | Mouse pachytene piRNA clusters are transcribed from canonical promoters. A-MYB drives their transcription bi-directionally or uni-directionally at the onset of meiosis. **e** | *C. elegans* type I 21U-RNAs are individually transcribed from mini-genes by the protein Forkhead (FKH), which binds the “*Ruby*” motif upstream of each piRNA precursor. The A-MYB-like transcription factor SNPC-4 — recruited by PRDE-1 — enhances transcription. Transcription of type II 21U-RNAs initiates at the promoters of full-length protein-coding genes.

FIG. 2 | Type I piRNA biogenesis in *C. elegans*. The initial precursor piRNA must be processed at both ends. Processing at the 5' end creates a monophosphorylated 5' end that can bind the PIWI protein PRG-1. The 3' end of the PIWI-bound pre-piRNA is then

trimmed by PARN-1 followed by 2'-O-methylation by HENN-1 to produce a mature 21U RNA.

FIG. 3 | piRNA-guided PIWI slicing (ping-pong pathway) generates a responder piRNA and initiates production of phased trailing piRNAs (phased pathway).

5 **FIG. 4** | **a** | piRNAs silence transposons transcriptionally and post-transcriptionally. Nuclear PIWI proteins are guided by piRNAs to nascent transposon transcripts and generate heterochromatin via DNA or histone methylation, silencing transcription. In the cytoplasm, piRNAs elicit post-transcriptional silencing by directing PIWI proteins to slice target transcripts. **b** | *C. elegans* piRNAs distinguish self from non-self transcripts. When
10 a PRG-1-bound 21U-RNA finds its target, it recruits RNA-dependent RNA polymerase (RdRP) to synthesize 22G-RNAs using the target as a transcription template. The 22G-RNAs are loaded into the worm-specific Argonautes, WAGOs, which silence non-self transcripts. In the nucleus, WAGO-9 silences non-self transcription by recruiting histone methyltransferases (HMT) and the HP1 homolog HPL2 to the target locus. Such RNA-
15 induced epigenetic silencing (RNAe) persists over generations. The Argonaute protein CSR-1 counteracts WAGO silencing, protecting self transcripts. The nuclear localization of CSR-1 suggests that CSR-1 may also license transcription of self transcripts. **c** | In some animals, somatic piRNAs fight viruses. When infected by a positive strand, single-stranded RNA virus, mosquitoes mount an antiviral piRNA-based response. Upon viral
20 replication, Piwi5 (loaded with 1U antisense piRNAs) and Ago3 (loaded with 10A sense piRNAs) participate in heterotypic ping-pong, consuming viral RNAs. **d** | In mouse male spermatocytes, pachytene piRNAs are first made as cells enter the pachytene stage of meiosis. Recent studies suggest a role for pachytene piRNAs in regulating gene

expression during meiosis and late spermiogenesis by directing PIWI proteins to cleave target mRNAs.

RNA-INDUCED SILENCING COMPLEX (RISC)

RISC is a multiprotein complex minimally comprising an Argonaute family protein and its RNA guide.

CANONICAL TRANSCRIPTION

- 5 This standard transcription requires a promoter (typically marked by H3K4me3) and generates RNA with a 5' 7-methylguanosine cap and a 3' poly(A) tail.

HETEROCHROMATIN

Heterochromatin is the tightly packed form of DNA whose histones are heavily modified with repressive marks, typically H3K9me3.

10 SPERMATOGONIA

Spermatogonia are the undifferentiated germ cells located at the periphery of seminiferous tubules. They undergo mitosis and later give rise to developing spermatocytes.

PACHYTENE

- 15 Stage of meiotic prophase I when homologous recombination occurs.

INITIATOR piRNA

piRNA which guides a PIWI protein to slice a piRNA precursor transcript, triggering production of responder and trailing piRNAs from it.

PRE-PRE-PIRNA

piRNA which guides a PIWI protein to slice a piRNA precursor transcript, triggering production of responder and trailing piRNAs from it.

RESPONDER PIRNA

- 5 piRNA whose 5' end is generated by initiator piRNA-guided PIWI-catalyzed slicing of a piRNA precursor transcript.

PRE-PIRNA

Intermediate product of piRNA biogenesis loaded into a PIWI protein. Pre-piRNAs are 3'-to-5' trimmed and 2'-O-methylated at their 3' termini to yield mature piRNAs.

10 TRAILING PRE-PIRNAS

A string of tail-to-head, "phased" trailing pre-piRNAs follows a responder piRNA. Both 5' and 3' ends of trailer piRNAs are produced by the stepwise endonucleolytic fragmentation of a piRNA precursor transcript.

K_{CAT}

- 15 In Michaelis-Menten enzyme kinetics, the catalytic constant k_{cat} represents the maximum number of substrate molecules converted to product per active site per unit time.

PIWI SLICER ACTIVITY

Endonucleolytic cleavage of the target RNA catalyzed by piRNA-guided PIWI proteins.

Table 1 | Proteins implicated in the piRNA pathway.

| <i>Drosophila melanogaster</i> | <i>M. musculus</i> | Lepidoptera | Function | Conservation | References |
|--------------------------------|--------------------|-------------|---|-------------------|--|
| Rhino (Rhi) | | | Promotes non-canonical transcription and suppresses splicing in dual-strand piRNA clusters | Fly-specific | {Andersen et al., 2017, #7306; Chen et al., 2016, #40928; Klattenhoff et al., 2009, #38065; Mohn et al., 2014, #77489; Zhang et al., 2014, #20677} |
| Deadlock (Del) | | | Collaborates with Rhino to promote transcription in dual-strand piRNA clusters | Fly-specific | {Andersen et al., 2017, #7306; Chen et al., 2016, #40928; Mohn et al., 2014, #77489} |
| Cutoff (Cuff) | | | Collaborates with Rhino to promote transcription and suppress termination in dual-strand piRNA clusters | Fly-specific | {Chen et al., 2016, #40928; Mohn et al., 2014, #77489; Pane et al., 2011, #61432} |
| Moonshiner (Moon) | | | Collaborates with Rhino and Del to promote transcription in dual-strand piRNA clusters | Fly-specific | {Andersen et al., 2017, #7306} |
| Cubitus interruptus (Ci) | GLI1/2/3 | Ci | Promotes transcription of uni-strand clusters in somatic follicle cells of fly gonads | Most animals | {Goriaux et al., 2014, #40466} |
| | MYBL1 (A-MYB) | | Promotes transcription of pachytene piRNA clusters | Birds and mammals | {Li et al., 2013, #8343} |
| Hel25E (aka UAP56) | DDX39B | Uap56 | Enables nuclear export of piRNA precursor transcripts to sites of piRNA production | Most animals | {Hur et al., 2016, #60320; Zhang et al., 2012, #59323} |

| | | | | | |
|---------------------|-----------------------|-------|---|--------------|--|
| Piwi | PIWIL4 (MIWI2) | | Nuclear PIWI protein that directs transcriptional silencing | Most animals | {Cox et al., 1998, #181; Cox et al., 2000, #367; Carmell et al., 2007, #88633; Darricarrère et al., 2013, #66914} |
| Asterix (Arx) | GTSE1 (aka CUE110) | Gtsf1 | Required for transcriptional silencing (flies) and piRNA ping- pong piRNA production (mice) | Most animals | {Muerdter et al., 2013, #86013; Ohtani et al., 2013, #99877; Dönertas et al., 2013, #77194; Yoshimura et al., 2009, #10843; Yoshimura et al., 2018, #76134} |
| Panoramix (Panx) | | | Required for transcriptional repression (flies) | Fly-specific | {Yu et al., 2015, #49989; Sienski et al., 2015, #44720} |
| Maelstrom (Mael) | MAEL | Mael | Required for transcriptional repression, piRNA production | Most animals | {Findley et al., 2003, #17179; Soper et al., 2008, #24253; Aravin et al., 2009, #72503; Sienski et al., 2012, #17594; Castaneda et al., 2014, #101799; Matsumoto et al., 2015, #14343} |

| | | | | | |
|----------------------|------------------------------------|-------------------|---|-----------------|--|
| Aubergine (Aub) | PIWIL1 (MIWI), PIWIL2 (MILI) | TnPiwi, Siwi | Cytoplasmic PIWI protein required for ping-pong piRNA production | Most animals | {Schupbach and Wieschaus, 1991, #33724; Aravin et al., 2004, #13479; Kuramochi- Miyagawa et al., 2001, #374; Deng and Lin, 2002, #69961; Kuramochi- Miyagawa et al., 2004, #77146; Kawaoka et al., 2009, #94517; Matsumoto et al., 2016, #4864} |
| Argonaute3 (Ago3) | | TnAgo3, BmAgo3 | Cytoplasmic PIWI protein required to initiate ping-pong and phased piRNA production | Insect-specific | {Gunawardane et al., 2007, #65138; Brennecke et al., 2007, #38790; Li et al., 2009, #70761; Kawaoka et al., 2009, #94517; Huang et al., 2014, #31127} |
| Zucchini (Zuc) | PLD6 | Zuc | Outer mitochondrial membrane protein, putative endonuclease required to establish the 3' ends of responder pre-piRNAs and both ends of phased, trailing pre- piRNAs | Most animals | {Schupbach and Wieschaus, 1991, #33724; Pane et al., 2007, #49901; Haase et al., 2010, #5392; Watanabe et al., 2011, #9704; Huang et al., 2011, #46149; Ipsaro et al., 2012, #70084; Nishimasu et al., 2012, #101700} |
| Minotaur (Mino) | GPAT2 | Mino | Required for phased piRNA biogenesis | Most animals | {Vagin et al., 2013, #40433; Shiromoto et al., 2013, #8818} |

| | | | | | |
|-------------------|---------|----------------|---|--------------|--|
| Gasz | GASZ | Gasz | Outer mitochondrial membrane protein required for phased piRNA biogenesis | Most animals | {Ma et al., 2009, #35135; Czech et al., 2013, #62448; Handler et al., 2013, #62852} |
| Armitage (Armi) | MOV10L1 | Armi | Helicase-motif protein required for phased piRNA biogenesis | Most animals | {Cook et al., 2004, #33495; Zheng et al., 2010, #61026; Frost et al., 2010, #78086; Zheng and Wang, 2012, #94859; Vourekas et al., 2015, #44878; Fu et al., 2016, #5688; Pandey et al., 2017, #44692} |
| Squash (Squ) | | | Piwi-interacting protein | Fly-specific | {Pane et al., 2007, #49901; Haase et al., 2010, #5392} |
| Vasa (Vas) | MVH | Vasa, BmVasa | Nuage component, RNA-dependent ATPase required for ping-pong piRNA biogenesis | Most animals | {Xiol et al., 2014, #98930; Nishida et al., 2015, #103232; Kuramochi-Miyagawa et al., 2010, #20775; Wenda et al., 2017, #94637; Fu et al., 2018, #18933} |
| Spindle-E (Spn-E) | TDRD9 | Spn-E, BmSpn-E | Nuage component, Tudor domain-containing protein, required for ping-pong piRNA biogenesis | Most animals | {Aravin et al., 2001, #228; Vagin et al., 2006, #48605; Lim and Kai, 2007, #84528; Aravin et al., 2009, #72503; Shoji et al., 2009, #21684; Wenda et al., 2017, #94637; Nishida et al., 2015, #103232} |

| | | | | | |
|--------------------|-------|------------|--|-----------------|--|
| Krimper (Krimp) | | Krimper | Nuage component, Tudor domain- containing protein that promotes heterotypic ping-pong (flies) | Insect-specific | {Lim and Kai, 2007, #84528; Webster et al., 2015, #37148; Sato et al., 2015, #29697} |
| Qin | RNF17 | Qin, BmQin | Nuage component, Tudor domain- containing protein that antagonizes homotypic ping-pong (flies) | Most animals | {Pan et al., 2005, #13674; Zhang et al., 2011, #79304; Anand and Kai, 2012, #12689; Zhang et al., 2014, #25621; Wasik et al., 2015, #84524; Nishida et al., 2015, #103232} |
| Tejas (Tej) | TDRD5 | Tejas | Nuage component, Tudor domain- containing protein | Most animals | {Smith et al., 2004, #56464; Patil and Kai, 2010, #101768; Yabuta et al., 2011, #73006; Ding et al., 2018, #65942} |
| Tapas | TDRD7 | | Nuage component, Tudor domain- containing protein | Most animals | {Hosokawa et al., 2007, #68973; Patil et al., 2014, #12442; Tanaka et al., 2011, #29526} |
| Vreteno (Vret) | TDRD1 | Vreteno | Nuage and Yb body component, Tudor domain-containing protein | Most animals | {Handler et al., 2011, #15938; Zamparini et al., 2011, #26439; Chuma et al., 2006, #46434; Reuter et al., 2009, #86589; Wang et al., 2009, #63594; Mathioudakis et al., 2012, #56828} |

| | | | | | |
|----------------------|-------------------|--------|---|--------------|--|
| Tudor | TDRD6 | Tudor | Tudor domain-containing protein | Most animals | {Vasileva et al., 2009, #98636; Nishida et al., 2009, #96922} |
| Sister of Yb (SoYb) | | | Tudor domain-containing protein, required for piRNA biogenesis (fly) | Fly-specific | {Handler et al., 2011, #15938} |
| Brother of Yb (BoYb) | TDRD12 | Tdrd12 | Tudor domain-containing protein, required for ping-pong piRNA biogenesis | Most animals | {Handler et al., 2011, #15938; Pandey et al., 2013, #17901; Yang et al., 2016, #30387} |
| fs(1)Yb (aka Yb) | | | Yb body component, Tudor domain-containing protein, required for piRNA production in somatic follicle cells | Fly-specific | {Szakmary et al., 2009, #24009; Saito et al., 2010, #47914; Olivieri et al., 2010, #35447} |
| Shutdown (Shu) | FKBP6 | Shu | Co-chaperone required to load guide into PIWI protein | Most animals | {Xiol et al., 2012, #37124; Preall et al., 2012, #93692; Olivieri et al., 2012, #61023} |
| Hsp83 | HSP90 | Hsp90 | Co-chaperone required to load guide into PIWI protein | Most animals | {Specchia et al., 2010, #26702; Xiol et al., 2012, #37124; Olivieri et al., 2012, #61023} |
| Nibbler (Nbr) | | | Pre-piRNA 3' trimming exonuclease | Fly-specific | {Han et al., 2011, #98377; Liu et al., 2011, #19273; Feltzin et al., 2015, #33454; Hayashi et al., 2016, #59204} |
| Papi | TDRKH (aka TDRD2) | Papi | Pre-piRNA 3' trimming co-factor | Most animals | {Liu et al., 2011, #36046; Honda et al., 2013, #59984; Saxe et al., 2013, #76672} |

| | | | | | |
|------|---|--------|---|--------------|--|
| | PNLDC1 (poly(A)- specific ribonuclease- like domain- containing 1) | PnlDc1 | Pre-piRNA 3' trimming exonuclease | Most animals | {Izumi et al., 2016, #4722; Tang et al., 2016, #19230; Ding et al., 2017, #33444; Zhang et al., 2017, #52582} |
| Hen1 | HENMT1 | Hen1 | SAM-dependent, methyltransferase required to modify 2' hydroxyl at piRNA 3' end | Most animals | {Saito et al., 2007, #28100; Horwich et al., 2007, #59920; Kirino and Mourelatos, 2007, #34897; Lim et al., 2015, #83471} |